Journal of Medical Virology 48:191–198 (1996)

Genotype Determination of Hepatitis C Virus From Northern India: Identification of a New Subtype

Aswini K. Panigrahi, Juan Roca, Subrat K. Acharya, Shahid Jameel, and Subrat K. Panda

Departments of Pathology (A.K.P., S.K.P.) and Gastroenterology (S.K.A.), All India Institute of Medical Sciences (AIIMS), and Virology Group (J.R., S.J.), International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India

Hepatitis C virus (HCV) shows substantial nucleotide sequence diversity distributed throughout the viral genome, with many variants showing only 68–79% overall sequence homology. This has led to problems in diagnosis of HCV using commercial immunoassays. Based on clustering of homologous sequences, various genotypes and subtypes of HCV have been described from different geographical regions. In the present study, 11 isolates from India were genotyped using sequence comparison for part of the nonstructural (NS5) and structural (core) regions. Parts of the genome covering 451 bp (nt 9–459) of the core gene and a 249 bp fragment (nt 7959-8207) of the NS5 gene were reverse transcribed and amplified using nested polymerase chain reaction (RT-PCR). The amplified fragments were cloned and sequenced. The classification into genotypes was done on the basis of phylogenetic analysis. Four isolates showed sequence homology to type 1b. Two of the isolates were classified as type 3a. One isolate was classified as type 3b and the remaining four isolates were found to be variants of type 3 but did not belong to any designated subtype. On the basis of phylogenetic analysis two of the unclassified isolates were put into a new subtype of 3 named as 3g. In one of these variants, parts of a 5'-noncoding (5' NCR; 204 bp), envelope-E1 (435 bp), and NS3 (502 bp) regions were also amplified, cloned, and sequenced. This study demonstrates the type 3 variants including a new subtype (3g) to be the major cause of HCV infection in India.

KEY WORDS: HCV, RT-PCR, phylogenetic analysis, genotype

INTRODUCTION

Hepatitis C virus (HCV) is a major cause of post-transfusion non-A, non-B hepatitis [Kuo et al., 1989; Choo et al., 1990]. It is associated with a significant proportion of chronic hepatitis, liver cirrhosis, and has been linked to hepatocellular carcinoma (HCC). About half of infected individuals develop chronic hepatitis, of whom 10–20% progress to liver cirrhosis with an increased risk of developing HCC. In the last 6 years, at least 16 HCV genomes have been sequenced in their entirety and information on genetic variation is available from different parts of the world.

HCV is a positive-stranded RNA virus, related to flavi- and pestiviruses. It has a 9.4 Kb single-stranded RNA genome [Choo et al., 1990] comprising a 5'-noncoding region (5'NCR) followed by a single continuous open reading frame (ORF) encoding a polypeptide of 3,010–3,011 amino acid residues and a 3'-noncoding region [3'NCR; Kato et al., 1990; Takamizawa et al., 1991]. Different isolates of HCV show substantial nucleotide sequence variation distributed throughout the viral genome [Okamoto et al., 1991, 1992]. Regions encoding the putative envelope proteins (E1, E2/NS1) are the most variable [Hijikata et al., 1991; Weiner et al., 1991] whereas the 5'NCR is the most conserved [Cha et al., 1991; Han et al., 1991; Bukh et al., 1992]. Comparison of HCV sequences has led to the identification of a number of distinct virus types that may differ from each other by as much as 33% over the whole viral genome [Choo et al., 1991; Okamoto et al., 1991, 1992; Chan et al., 1992; Mori et al., 1992]. Many workers investigating sequence variation among HCV isolates have proposed schemes of classification and nomenclature [Enomoto et al., 1990; Houghton et al., 1991; Cha et al., 1992; Chan et al., 1992; Mori et al., 1992; Okamoto et al., 1992; Bukh et al., 1993, 1994; Simmonds et al., 1993]. Recently, a proposal has been made for a unified system of nomenclature based on sequence analysis of at least one non-structural and one structural region [Simmonds et al., 1994a].

A large volume of information is available on the genetic diversity of HCV from different geographical

The nucleotide sequence data reported in this paper have been submitted to EMBL database and assigned the accession numbers X91297 to X91307 and X91416 to X91423.

Accepted for publication September 19, 1995.

Address reprint requests to Dr. S.K. Panda, Department of Pathology, AIIMS, Ansari Nagar, New Delhi-110 029, India.

TABLE I. Sequences and Sources of the Primers Used in This Study

Name	Position of 5′ base[a]	Polarity[b]	Sequences 5′ to 3′	Source
5′ NCR				
46	−295	+	CTGTGAGGAACTACTGTCTT	Okamoto et al. [1990]
317	10	−	GTGCTCATGGTGCACGGTCTACGAGACCTCCGG	Garson et al. [1990]
64	−276	+	TTCACGCAGAAAGCGTCTAG	Okamoto et al. [1990]
287	−26	−	CACTCGCAAGCACCCTATCAGGCATGCA	Garson et al. [1990]
NS5				
243	7904	+	TGGGGATCCCGTATGATACCCGCTGCTTTGA	Enomoto et al. [1990]
242	8304	−	GGCGGAATTCCTGGTCATAGCCTCCGTGAA	Enomoto et al. [1990]
554	7935	+	CTCAACCGTCACTGAACAGGACAT	Chan et al. [1992]
555	8227	−	CCACGACTAGATCATCTCCG	Chan et al. [1992]
Core				
CCS 261	−69	+	GAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGT	This study
CCS 805	515	−	CGACCGAATTCATTCCCTGTTGCATAGTTCACGCCGTCCT	This study
CCS 298	−32	+	AGTGCCCCGGGAGGTCTCGTAGACCGTGCACCATGAGCAC	This study
CCS 789	499	−	CTGTTGCATAGTTCACGCCGTCCTCCAGAACCCGGACACC	This study
Envelope (E1)				
CES 662	332	+	GGACCCCCGGCGTAGGTCGCGTAAT	This study
CES 1447	1142	−	CCAGCAAAAAGTAGCATCACAATCA	This study
CES 832	502	+	GGAATCTGCCCGGTTGCTCTTTTTCTATCT	This study
CES 1297	992	−	GTTGTGGGCGACCAGTTCATCATCA	This study
NS3				
CNS 4024	3694	+	CCACTGGCAGCGGTAAGAGCACAAA	This study
CNS 4069	3739	+	CTCAAGGGTATAAGGTGCTCGTACT	This study
CNS 4596	4291	−	CTATAGTTGGTATGACGGACACATC	This study

[a]Position of 5′ base according to HCV-1 sequence, A of polyprotein ATG taken as +1.
[b]Polarity: + sense, − antisense.

TABLE II. HCV Isolates Characterised in This Study*

Sequence no.	Patient	Diagnosis	NS5	Core
1	IND054	CAH	NA[a]	98.0− 1b
2	IND308	CAH	NA	94.3− 3b
3	IND1404	CAH	NA	96.6− 1b
4	IND101	CAH	91.9− 3a	NA
5	IND1061	Cirrhosis	91.4− 3a	NA
6	IND1346	Cirrhosis	85.6− 3b	NA
7	IND674	Cirrhosis	86.5− 3f	91.1− 3f
8	IND1192	Cirrhosis	95.5− 1b	96.0− 1b
9	IND1358	CAH	90.1− 3a	96.1− 1b
10	IND1452	CAH	80.6− 3b	90.2− 3b
11	IND1751	CAH	81.1− 3b	90.5− 3b

*Values represent maximum nucleotide percent homology to known HCV types.
[a]NA, not amplified; CAH, chronic active hepatitis.

areas, which has epidemiological and clinical significance. The HCV genotypes have distinct geographical distributions [Bukh et al., 1992; Cha et al., 1992; Chan et al., 1992; Okamoto et al., 1992; McOmish et al., 1994], determine the severity of liver disease, and present a prognostic marker for response to interferon therapy [Kanai et al., 1992; Hino et al., 1993]. Only limited information regarding HCV variants is available from developing countries, especially the Indian subcontinent [Bukh et al., 1993, 1994; Valliammai et al., 1995] and neighbouring regions, namely Thailand [Mori et al., 1992; Okamoto et al., 1993], Nepal [Tokita et al., 1994a], Vietnam [Tokita et al., 1994b], and Indonesia [Hotta et al., 1994; Okamoto et al., 1994].

In this study, we investigated HCV genotypes from Northern India and identified the prevailing genotypes based on nucleotide sequences from the core and NS5 regions. We also describe major amino acid changes in the core gene from a group of isolates which may be responsible for the atypical immunoreactivity [Panigrahi et al., 1994].

MATERIALS AND METHODS

Patients

The patients in this study were identified from a screening programme using second-generation immunoassays and were drawn from The All India Institute of Medical Sciences and G.B. Pant Hospitals, New Delhi, India. All the patients included in this study had clinically and histopathologically confirmed chronic liver disease. The patients were screened for HCV viraemia by reverse transcriptase nested-polymerase chain reaction (RT-PCR) based on the viral 5′NCR [Cha et al., 1991]. A group of HCV-RNA-positive samples (n = 11) were investigated for genotype determination. Informed consent was obtained from each patient prior to the study and institutional ethical clearance was obtained.

Amplification of HCV-RNA

Primers. The primers used for cDNA synthesis and PCR are listed in Table I. These were synthesised on an automated oligonucleotide synthesiser (Applied Biosystems, USA: Model 392) using phosphoramidite chemistry.

RNA preparation. RNA for RT-PCR was extracted from the serum by the guanidinium isothiocyanate-acid phenol method [Chomczynski and Sacchi, 1987]. Briefly, 100 μl of serum was treated with 500 μl of Solution D (4 M guanidinium isothiocyanate, 0.75 M sodium citrate, 0.5% Sarkosyl, 0.1 M 2-mercaptoethanol). After extraction with 500 μl of water-saturated (acidic) phenol and 200 μl of chloroform-isoamyl alcohol (49:1), RNA was precipitated from the aqueous phase, sequentially with isopropanol and ethanol. The pellets were washed twice with 70% ethanol and resuspended in 25 μl of diethylpyrocarbonate (DEPC)-treated water. Prior to RT-PCR, the RNA was denatured by heating at 95°C for 5 min, followed by rapid chilling on ice.

Amplification of RNA by PCR. For amplification of 5'NCR and NS5 regions RT-PCRs were carried out in 50 μl reaction volume with 25 pmol each of external primers (Table I), 1X Taq buffer, 0.05% W-1 (BRL, USA), 2.5 mM $MgCl_2$, 200 μM each dNTPs, 25 units of human placental ribonuclease inhibitor (HPRI; Amersham, UK), 4 units of AMV reverse transcriptase (Amersham), 1 unit of Taq DNA polymerase (BRL), and the RNA equivalent of 50 μl serum. RT and PCR steps were carried out in a single tube by programming the DNA thermocycler (Techne, UK) at 42°C for 1 hr followed by 94°C for 4 min and 35 cycling at 94°C for 1 min, 50°C for 1.3 min, and 73°C for 1.3 min. One tenth volume of the first PCR products were reamplified with internal primers for another 35 cycles at the same conditions. For amplification of core, envelope (E1), and NS3, HCV-RNA was reverse transcribed into cDNA in the presence of 50 pmol of the external antisense primer (Table I) using 8 units of AMV reverse transcriptase (Amersham) in a 20 μl reaction volume at 42°C for 1 hr. A nested set of amplification was carried out separately as above using the following conditions: core, 94°C for 1 min, 72°C for 2 min; envelope (E1) and NS3, 94°C for 1 min, 55°C for 1 min, 73°C for 1.3 min. For NS3 amplification the external antisense primer was used in the nested step as the antisense primer. The size of the products from the 5'NCR, core, envelope, NS3, and NS5 were 251, 531, 481, 552, and 294 bp, respectively.

Cloning and Sequencing of PCR Products

PCR products were purified from agarose gels using the Gene Clean kit (Bio-101, USA) and individual fragments were cloned into the pCR-Script (SK+) vector (Stratagene, GmBH) as per manufacturers' guidelines. Multiple independent clones (at least three) were sequenced in both directions using T7 DNA polymerase (Sequenase ver 2.0, USB,USA) with T3, KS, and M13 primers. The consensus nucleotide sequences for each individual isolate were determined and used for further analysis.

Phylogenetic Analysis

Nucleotide sequences were aligned with CLUSTAL using the PC/Gene software. The sequences which were observed to be variants were compared to the EMBL database using the BLAST alignment search programme. Phylogenetic analyses were carried out with the SEQBOOT, DNADIST, and FITCH algorithms in the PHYLIP ver 3.41 package [Macintosh version; Felsenstein, 1989].

RESULTS

Comparison of Sequences in the NS5 Region

The NS5 region (nt 7959-8207) was amplified in 8 of 11 isolates (Table II). The nucleotide sequences of the region nt 7975-8196 (222 bp) from these eight cases were compared with sequences available in the EMBL database. Three isolates (IND101, IND1061, and IND1358) showed more than 90% homology with genotype 3a (Table II) and were 96% homologous between themselves. One isolate (IND1192) showed 95.5% sequence homology with genotype 1b (Table II). One isolate (IND674) showed 86.5% nucleotide sequence homology with a Nepalese strain NE125 [Tokita et al., 1994a] belonging to subtype 3f (Table III). Two isolates (IND1452 and IND1751) showed a maximum sequence homology of 71–81% with any of the published sequences belonging to type 3 (Table III). These two isolates showed a sequence homology of 97.7% amongst themselves. Isolate IND1346 had a maximum sequence homology of 85.6% with type 3b sequences and 88.3% with the isolate IND1751 sequence (Table III). Therefore, isolates IND1452, IND1346, and IND1751 did not fall into any known subtype.

Comparison of Sequences in the Core Region

The core gene from nucleotide position 9 to 459 was RT-PCR amplified and sequenced from 8 of 11 isolates (Table II). Four of the isolates (IND054, IND1192, IND1358, and IND1404) showed greater than 96% nucleotide sequence homology to HCV type 1b (Table II). One isolate (IND308) showed 94.3% sequence homology to the Nepalese isolate NE137 [Tokita et al., 1994a] belonging to type 3b (Table III). Two other isolates (IND1452 and IND1751) have a maximum sequence homology of 90.5% with type 3b sequences and were 99% homologous to each other (Table III). The isolate (IND674) had a maximum homology of 91.4% with isolates IND1452 and IND1751 and 91.1% with the Nepalese isolate NE125 [Tokita et al., 1994a] belonging to subtype 3f (Table III).

Phylogenetic Relationship

Based on the nucleotide sequence comparisons (Table II), the isolates IND054, IND1192, and IND1404 are classified as genotype 1b. The isolate IND1358 showed maximum sequence homology in the core region of

TABLE III. Pairwise Comparisons of Partial Nucleotide Sequences in the NS5 and Core Regions of HCV Isolates Belonging to Genotype 3*

	3a									3b						3c	3d	3e	3f		?	?		3g		?
	T1	T7	TH85	14314	T90	EB7	S21	IND1061	IND101	T9	T10	NE137	MN6	MN7	IND308	NE048	NE274	NE145	NE125	A21	IND674	TD3	MN5	IND1751	IND1452	IND1346
															NS5											
T1	100	94.6	95.5	95.0	95.6	95.0	96.8	89.6	90.1	76.6	77.5	77.5	77.0	75.7	–	83.3	81.5	79.7	77.5	78.0	77.5	72.6	71.6	76.6	76.1	80.6
T7		100	97.7	96.4	92.9	93.7	93.7	89.6	89.2	78.9	79.2	78.4	79.2	77.0	–	83.3	80.6	80.6	78.4	78.9	76.6	71.6	69.8	76.6	76.1	81.1
TH85			100	98.6	95.1	95.6	95.9	90.1	89.6	77.5	78.0	77.0	77.0	75.7	–	81.5	81.1	81.0	79.2	79.7	77.5	72.1	70.3	77.5	77.0	80.2
14314				100	94.7	95.1	95.5	90.1	89.2	78.0	78.4	77.0	77.5	76.1	–	81.5	80.6	80.6	79.3	79.8	77.0	73.0	71.2	77.0	76.6	79.8
T90					100	96.8	96.0	88.7	89.2	77.5	78.0	78.0	77.5	76.6	–	81.5	82.0	79.7	78.8	79.3	78.0	73.0	71.2	74.7	74.3	78.4
EB7						100	95.5	88.7	90.1	78.0	78.4	78.4	78.4	77.0	–	82.4	82.9	80.6	78.0	78.4	78.4	74.7	73.0	75.2	75.2	78.9
S21							100	91.4	91.9	76.6	77.0	76.1	76.1	75.7	–	82.0	81.5	81.0	78.4	78.9	78.4	73.0	72.1	76.6	76.1	79.2
IND1061								100	96.8	78.0	78.4	73.8	78.0	76.6	–	81.5	77.0	80.6	77.0	77.5	75.2	70.7	70.7	73.5	73.0	76.1
IND101									100	76.6	77.0	74.3	76.1	74.7	–	82.0	78.4	81.0	78.9	79.2	77.0	71.2	72.1	73.9	73.8	77.0
T9										100	99.1	91.9	93.3	94.6	–	78.0	76.1	74.3	82.0	81.5	78.8	72.6	72.1	80.1	79.7	85.6
T10											100	91.5	93.2	94.6	–	78.4	75.7	74.7	82.4	82.0	79.7	72.1	71.6	81.1	80.6	85.6
NE137			88.0									100	89.6	93.7	–	78.9	75.7	74.7	79.7	79.2	77.5	71.2	70.3	79.7	79.2	81.5
MN6													100	94.1	–	77.5	75.2	73.8	79.7	79.2	78.4	73.9	71.6	78.4	77.9	82.4
MN7														100	–	77.5	74.7	74.3	79.7	79.2	78.8	71.2	70.7	81.1	80.6	83.3
IND308			87.1									94.3			100	–	–	–	–	–	–	–	–	–	–	–
NE048			89.4									88.7			88.5	100	81.5	82.0	77.5	77.0	73.8	74.7	73.8	75.7	75.2	79.7
NE274			89.6									87.6			87.6	87.6	100	79.2	79.2	79.7	79.7	74.7	73.0	75.7	75.7	80.6
NE145			89.6									86.3			85.1	87.8	87.4	100	77.0	76.6	76.6	67.1	70.3	72.1	71.2	73.0
NE125			87.6									91.8			89.4	87.4	88.0	86.2	100	99.5	86.5	71.6	72.5	81.0	81.0	81.5
A21																				100	86.9	72.1	73.0	80.6	80.6	81.0
IND674			87.1									90.2			89.4	86.2	85.6	86.2	91.1		100	73.0	74.4	80.2	80.2	79.7
TD3																						100	92.3	72.1	72.1	73.5
MN5																							100	71.2	71.2	73.9
IND1751			85.8									90.5			89.1	87.6	85.4	84.5	90.2		91.4			100	97.7	88.3
IND1452			85.6									90.2			88.9	87.4	85.1	84.2	90.0		91.4			99.1	100	87.8
IND1346																										100
															Core											

*Numbers represent percentage nucleotide identity. Blank space represents data not available.

96.2% to type 1b, but in the NS5 region of 90.1% to type 3a, indicating coinfection. The three clones sequenced from each of the regions from this isolate (IND1358) showed complete sequence homology amongst themselves. The isolate IND308 belonged to type 3b and isolates IND674, IND1346, IND1452, and IND1751 belonged to type 3 but could not be classified to any existing subtype. Comparison of these sequences to the other type 3 isolates is described in Table III. We carried out a phylogenetic comparison of core and NS5 regions of isolates IND308, IND674, IND1346, IND1452, and IND1751 with all subtypes of type 3 so far reported. The resulting DNA distance trees showed similar branching patterns into subtypes (Fig. 1) except for isolate IND674, which clustered with isolated IND1452 and IND1751 in case of core sequences and NE125 (subtype 3f) in case of NS5 sequences. Isolates IND1452 and IND1751 were present as a separate cluster altogether in both core and NS5 trees. The isolate IND1346, although placed on the same branch of the phylogenetic tree showed some difference from the other isolates (IND1452 and IND1751).

Sequences of 5'NCR, Envelope (E1), and NS3 Region of Isolate IND1751

Table IV compares the nucleotide sequence of 5'NCR (nt -256 to -53), env(E1) (nt 533-967), and NS3 (nt 3764-4265) regions from the isolate IND1751 with the major genotypes of HCV. The maximum sequence homology observed in the 5'NCR was 96.6% with subtype 3e [NE145, Tokita et al., 1994a] followed by 96% with subtypes 3b and 3a. The homology with other genotypes including 7, 8, and 9 [Tokita et al., 1994b] ranged between 88.2% and 91.7%. In the envelope (E1) region, the homology was in the range of 56.8–75.9% and in the NS3 region it was 69.8–76.9%.

Amino Acid Sequence Comparison

The predicted amino acid sequences from the core, envelope (E1), NS3, and NS5 of isolate IND1751 were compared with the major genotypes of HCV (Table IV). Amino acid sequence of the core region of this variant when compared with the major HCV genotypes showed nonconservative changes at positions 10, 16, 21, 50, 63, 64, 70,71, 142, and 150 (Fig. 2).

DISCUSSION

HCV is characterised by a high degree of nucleotide sequence heterogeneity, with sequences in the 5'NCR being highly conserved, with a 92–98% identity [Okamoto et al., 1992]. Among the putative coding regions, the core gene is highly conserved with a nucleotide sequence identity of 81–92%, whereas the envelope (E1) and E2/NS1 are highly variable with nucleotide sequence identities of only 55–75%. The nonstructural genes, NS2 to NS5, also show variation with a 57–79% nucleotide sequence homology. Of the several schemes proposed initially for classification and nomenclature of HCV genotypes [Enomoto et al., 1990; Houghton et al., 1991; Cha et al., 1992; Chan et al., 1992; Mori et al., 1992; Okamoto et al., 1992; Bukh et al., 1993, 1994; Simmonds et al., 1993], the proposal made by Simmonds et al. [1994a] for a unified system of nomenclature has now been adopted. This system recognises that the most divergent isolates of HCV have sequence similarities of less than 72% that are classified as types. Within these groupings (types), there may be only 75–86% sequence similarity between clusters of isolates which are classified as subtypes.

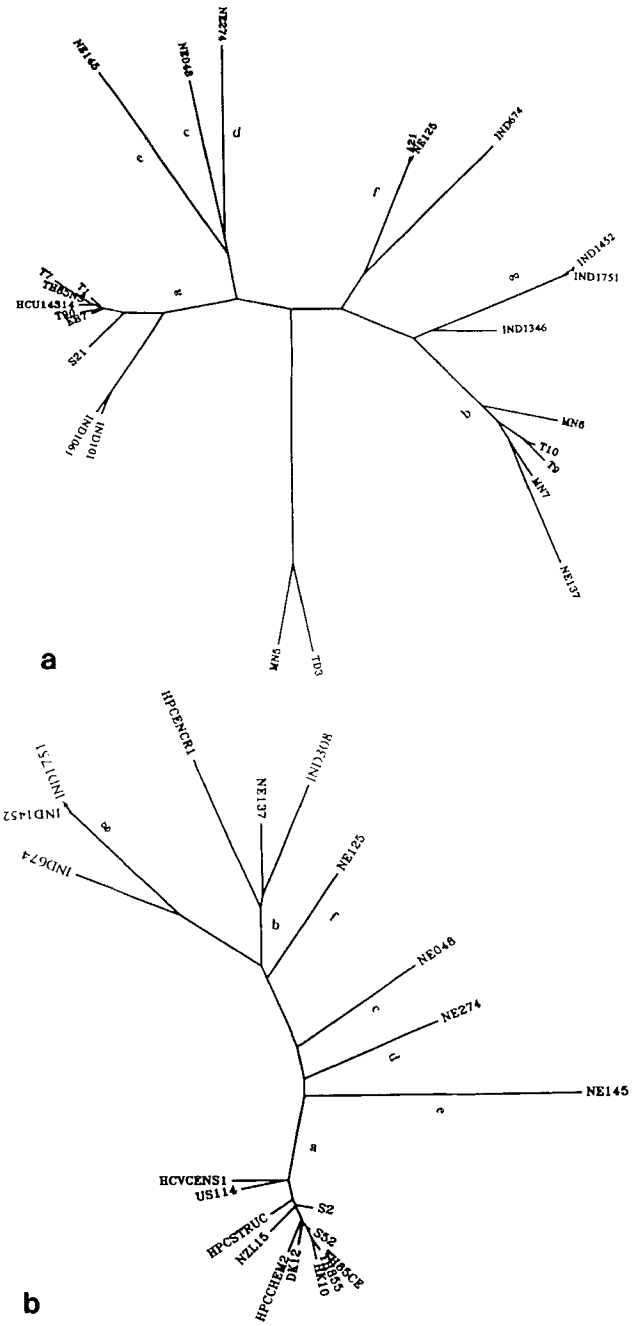

Fig. 1. Phylogenetic analysis of type 3 HCV isolates. **a:** NS5 sequences (7975–8196 nt) from 25 isolates showing eight subtypes including a new subtype designated as 3g. **b:** Core sequences (9-459 nt) from 25 isolates, confirming the classification into a new subtype.

TABLE IV. Sequence Homology* of HCV Type 3g With Major Genotypes

Region	1a	1b	2a	2b	3a	3b
5'NCR	91.2	91.7	88.2	85.8	96.1	96.1
Core	82.7	83.1	78.7	78.5	86.0	90.5
	(86.3)	(86.3)	(83.7)	(83.0)	(91.5)	(93.5)
Env (E1)	63.1	63.7	60.0	56.8	75.2	75.9
	(70.1)	(66.7)	(58.3)	(58.3)	(80.6)	(80.6)
NS3	71.6	69.8	68.1	70.5	76.9	—
	(82.6)	(81.4)	(80.8)	(80.8)	(88.6)	
NS5	66.7	65.5	62.2	62.7	79.1	81.9
	(72.0)	(73.2)	(68.3)	(72.0)	(86.6)	(84.1)

*Numbers represent percent nucleotide homology. Values in parentheses represent percent amino acid homology.

In the present study, the method proposed by Simmonds et al. [1994a] was used for classification of HCV isolates from Northern India. A phylogenetic analysis was carried out for all the isolates using core and NS5 region sequences separately, along with similar sequences from published prototypes of each genotype and subtype. Both the core and NS5 trees showed similar branching patterns with isolates IND1751, IND1452, IND1346, and IND674 and formed a separate cluster on the type 3a branch (results not shown). Three of 11 isolates (27.3%) were of type 1b and 2 of 11 (18.2%) were of type 3a. In one case the core sequence showed maximum homology with type 1b but the NS5 sequence showed homology with type 3a. In this study, we also observed that the NS5 region could not be amplified in most of the type 1b cases in which the core was easily amplifiable. On the other hand, in type 3a the core region was difficult to amplify but the NS5 region was easily amplifiable (Table II). The use of type-specific primers may help in resolving this issue. All variants that were assigned a particular genotype in one region of the viral genome always contained a sequence corresponding to the same genotype elsewhere in the genome [Simmonds et al., 1994b]. Based on this observation, it is possible that this case (IND1358) had a coinfection with HCV types 1b and 3a. The five other cases (45.5%) were observed to be variants of type 3a. Of these, one case (IND308) was classified as 3b based on the core sequence alone as the NS5 region could not be amplified from this case (Fig. 1b).

Recently, Valliammai et al. [1995] have isolated a new subtype of HCV from the Southern part of India (isolate MN5) that showed sequence homology to an Indonesian isolate, TD-3 [Hotta et al., 1994]. Both these isolates were described to be a new subtype of HCV type 3 [Valliammai et al., 1995]. Isolates TD-3 and MN5 are 70.7% homologous with IND1061 (Table III), indicating that they belong to two different genotypes. The degree of homology is comparable with those between type 8 and 9 isolates. Phylogenetic analysis of isolate TD-3 with all HCV types showed that this isolate belongs to a new genotype (results not shown), which may be designated as type 10a instead of a subtype of type 3 as described by Valliamai et al. [1995]. Two of the isolates (IND1452 and IND1751) formed a separate cluster on a phylogenetic tree of all type 3 sequences from both the NS5 and core regions (Fig. 1). These isolates may be classified as a new subtype. One other isolate, IND1346, in which only the NS5 sequence was available, may belong to this group based on nucleotide sequence homology. However, this isolate may belong to another type 3 subtype based on phylogeny (Fig. 1a). Another unclassified isolate, IND674, branched with subtype 3f [Tokita et al., 1994a] in the NS5 region whereas isolates IND1452 and IND1751 branched in the core region (Fig. 1). This serum sample was from a case of liver cirrhosis with persistent viraemia. The possibility of active mutation leading to this variation cannot be completely ignored. However, it is more likely that this isolate belongs to yet another subtype of the type 3 group.

The sequence data of the NS5 (non-structural) region are supplemented with those of the core region (structural) for isolates IND1452 and IND1751. As phylogenetic analysis correlates in both regions, these two isolates may formally be classified as HCV type 3g according to recommendations of Simmonds et al. [1994a]. To confirm further this classification, we generated sequence data from 5'NCR, env(E1), and NS3 regions in one of these isolates (IND1751). The sequences were compared with the existing major genotypes (Table IV) and significant variation was observed, thereby further substantiating the grouping of these isolates into a new subtype.

The predicted amino acid sequences of the subtype 3g were compared with those of the major types of HCV (Table IV). In the core region, homology ranged from 83% with type 2b to a maximum of 93.5% with 3b. Similarly, in the envelope (E1) region 58.3–80.6%, in the NS3 region 80.8–88.6%, and in the NS5 region 68.3–86.6% homology was observed. The core region of this new subtype 3g showed amino acid changes unique to it (Fig. 2). Of these, the non-conservative changes that may lead to an altered conformation are $K_{10} \rightarrow Q$, $N_{16} \rightarrow P$, $D_{21} \rightarrow N$, $R_{50} \rightarrow V$, $P_{64} \rightarrow V$, $A_{142} \rightarrow P$, and $A_{150} \rightarrow R$. Three of these mutations, two of which lead to an altered charge ($K_{10} \rightarrow Q$, $D_{21} \rightarrow N$) and the other to a turn ($N_{16} \rightarrow P$), were concentrated in the N-terminal immunodominant portion of the core polypeptide [Ishida et al., 1993]. Such changes may prove important in the differential immunodiagnosis of subtype 3g cases and its pathogenicity.

We observed that HCV variants of genotype 3 predominate in Northern India followed by type 1b. It is important to note that from neighbouring countries such as Nepal and Thailand, HCV type 3 variants have been described. However, the isolates from southern India resemble more closely the sequences from Indonesia. This follows the pattern of population migration and trade routes of two parts of the Indian subcontinent. A new subtype of HCV from this region, provisionally designated 3g, is described. This study also confirmed that HCV is very heterogeneous, even among isolates belonging to the same genotypes. We are further characterising this subtype and developing a

```
1a   MSTNPKPQKK NKRNTNRRPQ DVKFPGGGQI VGGVYLLPRR GPRLGVRATR
1b   --------R- T--------- ---------- ---------- ----------
2a   --------R- T--------- ---------- ---------- ----------
2b   --------R- T--------- ---------- ---------- ----------
3a   ---L----R- T----I---- ---------- -----V---- ----------
3b   ---L----RQ T----Y---- N--------- -----V---- ---------V
3g   ---L----RQ T----P---- N--------- -----V---- ---------V

1a   KTSERSQPRG RRQPIPKARR PEGRTWAQPG YPWPLYGNEG CGWAGWLLSP
1b   ---------- ---------Q ----A----- ---------- L---------
2a   ---------- -------D-- ST-KS-GK-- ---------- L---------
2b   ---------- -------D-- ST-KS-GK-- ---------- ----------
3a   ---------- ---------- S---S----- ---------- ----------
3b   ---------- ---------S R---S----- ---------- ----------
3g   ---------- --HV-----Q R---S----- ---------- ----------

1a   RGSRPSWGPT DPRRRSRNLG KVIDTLTCGF ADLMGYIPLV GAPLGGAARA LAH
1b   Y----R---- ---------- ---------- ---------- ---------- ---
2a   ---------N ---H----V- ---------- --------V- ------V--- ---
2b   -----T---- ---H------ R----I---- --------V- ---V--V--- ---
3a   ---------N ---------- ---------- ---------- ---V--V--- ---
3b   ---------N ---------- ---------- ---------I ---V--V--- ---
3g   ---------N ---------- ---------- ---L--V--- -P-V--V--R F--
```

Fig. 2. Amino acid sequence comparison of subtype 3g core region (1-153 aa) with major HCV types. Dashes represent conserved amino acids.

type-specific immunoassay to detect it in the hope of reducing false-negative results [Panigrahi et al., 1994]. The pathogenicity and response to interferon therapy of this new subtype remain interesting areas for future studies.

ACKNOWLEDGMENTS

This study was primarily supported by a grant from the Indian Council of Medical Research (ICMR), New Delhi, India. We thank Dr. S.K. Sarin of the G.B. Pant Hospital, New Delhi, for sera from HCV patients. A.K. Panigrahi is a recipient of a CSIR senior research fellowship, New Delhi. Juan Roca is the recipient of an ICGEB Training Fellowship. This work was partially supported by internal funds of ICGEB, New Delhi.

REFERENCES

Bukh J, Purcell RH, Miller RH (1992): Sequence analysis of 5′ noncoding region of hepatitis C virus. Proceedings of the National Academy of Sciences, USA 89:4942–4946.

Bukh J, Purcell RH, Miller RH (1993): At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proceedings of the National Academy of Sciences, USA 90:8234–8238.

Bukh J, Purcell RH, Miller RH (1994): Sequence analysis of core gene of 14 hepatitis C virus genotypes. Proceedings of National Academy of Sciences, USA 91:8239–8243.

Cha TA, Kolberg J, Irvine B, Stempien M, Beall E, Yano M, Choo Q-L, Houghton M, Kuo G, Han JH, Urdea MS (1991): Use of a signature nucleotide sequence of hepatitis C virus for detection of viral RNA in human serum and plasma. Journal of Clinical Microbiology 29:2528–2534.

Cha TA, Beall E, Irvine B, Kolberg J, Chien D, Kuo G, Urdea MS (1992): At least five related, but distinct, hepatitis C viral genomes exist. Proceedings of National Academy of Sciences, USA 89:7144–7148.

Chan S-W, McOmish F, Holmes EC, Dow B, Peutherer JF, Follett E, Yap PL, Simmonds P (1992): Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants. Journal of General Virology 73:1131–1141.

Chomczynski P, Sacchi N (1987): Single step method of RNA isolation by guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry 162:152–159.

Choo Q-L, Weiner AJ, Overby LR, Kuo G, Houghton M, Bradley DW (1990): Hepatitis C virus—The major causative agent of viral non-A, non-B hepatitis. British Medical Bulletin 46:423–441.

Chool Q-L, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina-Selby R, Barr PJ, Weiner AJ, Bradley DW, Kuo G, Houghton M (1991): Genetic organization and diversity of the hepatitis C virus. Proceedings of the National Academy of Sciences, USA 88:2451–2455.

Enomoto N, Takada A, Nakao T, Date T (1990): There are two major types of hepatitis C virus in Japan. Biochemical and Biophysical Research Communications 170:1021–1025.

Felsenstein J (1989): PHYLIP—Phylogeny inference package (version 3.2). Cladistics 5:164–166.

Garson JA, Ring C, Tuke P, Tedder RS (1990): Enhanced detection by PCR of hepatitis C virus RNA. Lancet 336:878–879.

Han JH, Shyamala V, Richman KH, Brauer MJ, Irvine B, Urdea MS, Tekampolson P, Kuo G, Choo Q-L, Houghton M (1991): Characterization of the terminal regions of hepatitis C viral RNA—Identification of conserved sequences in the 5′ untranslated region and poly(A) tails at the 3′ end. Proceedings of the National Academy of Sciences, USA 88:1711–1715.

Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Ohkoshi S, Shimotohno K (1991): Hypervariable region in the putative glycoprotein of hepatitis C virus. Biochemical and Biophysical Research Communications 175:220–228.

Hino K, Sainokami S, Shimoda K, Iino S, Wang Y, Okamoto H, Miyakawa Y, Mayumi M (1993): Genotypes and titers of hepatitis C virus for predicting response to interferon in patients with chronic hepatitis C. Journal of Medical Virology 42:299–305.

Hotta H, Handajani R, Lusida MI, Soemarto W, Doi H, Miyajima H, Homma M (1994): Subtype analysis of hepatitis C virus in Indonesia on the basis of NS5b region sequences. Journal of Clinical Microbiology 32:3049–3051.

Houghton M, Weiner A, Han J, Kuo G, Choo Q-L (1991): Molecular biology of hepatitis C viruses: Implications for diagnosis, development and control of viral disease. Hepatology 14:381–388.

Ishida C, Matsumoto K, Fukada K, Matsushita K, Shiraki H, Maeda Y (1993): Detection of antibodies to hepatitis C virus (HCV) structural proteins in anti-HCV-positive sera by an enzyme-linked immunosorbent assay using synthetic peptides as antigens. Journal of Clinical Microbiology 31:936–940.

Kanai K, Kako M, Okamoto H (1992): HCV genotypes in chronic hepatitis C and response to inferferon. Lancet 339:1543.

Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, Shimotonho K (1990): Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proceedings of the National Academy of Sciences, USA 87:9524–9528.

Kuo G, Choo Q-L, Alter HJ, Gitnick GL, Redekar AG, Purcell RH, Miyamura T, Dienstag JL, Alter MJ, Stevens CE (1989): An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science 244:362–364.

McOmish F, Yap PL, Dow BC, Follett EAC, Seed C, Keller AJ, Cobain TJ, Krusius T, Kolho E, Naukkarinen R, Lin C, Lal C, Leong S, Medgyesi GA, Hejja M, Kiyokawa H, Fukuda K, Cuypers T, Saeed AA, Al-Rasheed AM, Lin M, Simmonds P (1994): Geographic distribution of hepatitis C virus genotypes in blood donors: An international colloaborative survery. Journal of Clinical Microbiology 32:884–892.

Mori S, Kato N, Yagyu A, Tanaka T, Ikeda Y, Petchclai B, Chewsilp P, Kurimura T, Shimotohno K (1992): A new type of hepatitis C virus in patients in Thailand. Biochemical and Biophysical Research Communication 183:334–342.

Okamoto H, Okada S, Sugiyama Y, Yotsumoto S, Tanaka T, Yoshizawa H, Tsuda F, Miyakawa Y, Mayumi M (1990): The 5′-terminal sequence of the hepatitis C virus genome. Japanese Journal of Experimental Medicine 60:167–177.

Okamoto H, Okada S, Sugiyama Y, Kurai K, Iizuka H, Machida A, Miyakawa Y, Mayumi M (1991); Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: Comparison with reported isolates for conserved and divergent regions. Journal of General Virology 72:2697–2704.

Okamoto H, Kurai K, Okada S, Yamamoto K, Iizuka H, Tanaka T, Fukuda S, Tsuda F, Mishiro S (1992): Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: Comparative study of four distinct genotypes. Virology 188: 331–341.

Okamoto H, Tokita H, Sakamoto M, Horikita M, Kojima M, Iizuka H, Mishiro S (1993): Characterization of the genomic sequence of type V (or 3a) hepatitis C virus isolates and PCR primers for specific detection. Journal of General Virology 74:2385–2390.

Okamoto H, Kojima M, Sakamoto M, Iizuka H, Hadiwandowo S, Suwignyo S, Miyakawa Y, Mayumi M (1994): The entire nucleotide sequence and classification of hepatitis C virus isolate of a novel genotype from an Indonesian patient with chronic liver disease. Journal of General Virology 75:629–635.

Panigrahi AK, Nanda SK, Dixit RK, Acharya SK, Zuckerman AJ, Panda SK (1994): Diagnosis of hepatitis C virus-associated chronic liver disease in India: Comparison of HCV antibody assay with a polymerase chain reaction for the 5′ noncoding region. Journal of Medical Virology 44:176–179.

Simmonds P, Holmes EC, Cha T-A, Chan S-W, McOmish F, Irvine B, Beall E, Yap PL, Kolberg J, Urdea MS (1993): Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. Journal of General Virology 74:2391–2399.

Simmonds P, Alberti A, Alter H, Bonino F, Bradley DW, Brechot C, Brouwer JT, Chan S-W, Chayama K, Chen D-S, Choo Q-L, Colombo M, Cuypers HTM, Date T, Dusheiko GM, Esteban JI, Fay O, Hadziyannis SJ, Han J, Hatzakis A, Holmes EC, Hotta H, Houghton M, Irvine B, Kohara M, Kolberg JA, Kuo G, Lau JYN, Lelie PN, Maertens G, McOmish F, Miyamura T, Mizokami M, Nomoto A, Prince AM, Reesink HW, Rice CM, Roggendorf M, Schalm SW, Shikata T, Shimotohno K, Stuyver L, Trepo C, Weiner A, Yap PL, Urdea MS (1994a); A proposed system for the nomenclature of hepatitis C viral genotypes. Hepatology 19:1321–1324.

Simmonds P, Smith DB, McOmish F, Yap PL, Kolberg J, Urdea MS, Holmes EC (1994b): Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. Journal of General Virology 75:1053–1061.

Takamizawa A, Mori C, Fuke I, Manabe S, Murukami S, Fujita J, Onishi E, Andoh T, Yoshida I, Okayama H (1991): Structure and organization of the hepatitis C virus genome isolated from human carriers. Journal of Virology 65:1105–1113.

Tokita H, Shrestha SM, Okamoto H, Sakamoto M, Horikita M, Iizuka H, Shrestha S, Miyakawa Y, Mayumi M (1994a): Hepatitis C virus variants from Nepal with novel genotypes and their classification into the third major group. Journal of General Virology 75:931–936.

Tokita H, Okamoto H, Tsuda F, Song P, Nakata S, Chosa T, Iizuka H, Mishiro S, Miyakawa Y, Mayumi M (1994b): Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth and ninth major genetic groups. Proceedings of the National Academy of Sciences, USA 91:11022–11026.

Valliammai T, Thyagarajan SP, Zuckerman AJ, Harrison TJ (1995): Diversity of genotypes of hepatitis C virus in southern India. Journal of General Virology 76:711–716.

Weiner AJ, Brauer MJ, Rosenblatt J, Richman KH, Tung J, Crawford K, Bonino F, Saracco G, Choo Q-L, Houghton M, Han JH (1991): Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology 180:842–848.